

Table S1. Clinical and epidemiological characteristics of sporadic breast cancer patients (SBC) and healthy controls (C) analyzed on 450K HumanMethylation BeadChip platform (discovery cohort). BMI: body mass index, N/A: not applicable.

Variable	Breast cancer patients	Healthy controls	P value
Age at diagnosis/recruitment (years)			
mean \pm SD	59 \pm 9.2	62.6 \pm 7.9	0.293
BMI (kg/m²)			
mean \pm SD	29.4 \pm 4.9	28.6 \pm 5.6	0.716
Genetic ancestry (%) mean \pm SD			
European	76.4 \pm 11.5	79.0 \pm 12.8	0.580
African	9.4 \pm 7.7	6.6 \pm 6.9	0.351
Native american	14.2 \pm 11.1	14.4 \pm 9.7	0.967
Smoking (%)			
Yes	31.8	0	1
No	68.2	100	
Tumor stage (%)			
I	26.3	N/A	
II	57.9	N/A	
III	15.8	N/A	
IV	0	N/A	
Histological type (%)			
Ductal	84.2	N/A	
Lobular	10.5	N/A	
Papillary	5.3	N/A	
Estrogen receptor (%)			
Positive	77.8	N/A	
Negative	22.2	N/A	
Progesterone receptor (%)			
Positive	66.7	N/A	
Negative	33.3	N/A	
Her2 (%)			
Positive	38.9	N/A	
Negative	61.1	N/A	

Table S2. Clinical and epidemiological characteristics of sporadic breast cancer patients and healthy controls used as an independent sample to validate candidate CpG sites (validation cohort). N/A: not applicable.

Variable	Breast cancer patients	Healthy controls	P value
Age at diagnosis/recruitment (years)			
mean \pm SD	55.9 \pm 10.0	53.6 \pm 9.7	0.147
BMI (kg/m²)			
mean \pm SD	27.7 \pm 5.7	26.5 \pm 5.4	0.176
Genetic ancestry (%) mean \pm SD			
European	78.4 \pm 12.5	76.0 \pm 14.1	0.258
African	9.3 \pm 7.5	9.4 \pm 7.7	0.955
Native american	12.3 \pm 9.3	14.6 \pm 10.9	0.150
Smoking (%)			
Yes	53.4	51.3	0.815
No	46.6	48.7	
Tumor stage (%)			
I	34.7	N/A	
II	34.6	N/A	
III	19.2	N/A	
IV	11.5	N/A	
Histological type (%)			
Ductal	84.0	N/A	
Lobular	14.0	N/A	
Papillary	2.0	N/A	
Estrogen receptor (%)			
Positive	66.0	N/A	
Negative	34.0	N/A	
Progesterone receptor (%)			
Positive	66.7	N/A	
Negative	33.3	N/A	
Her2 (%)			
Positive	28.9	N/A	
Negative	71.1	N/A	

Table S3. Clinical characteristic of tumors corresponding to paired breast tissues (Normal/Tumor) of independent European cohort of breast cancer patients. TNP: triple negative phenotype, N: normal breast tissue, T: breast tumor tissue, NA: data not available.

Paired tissues	Subtype	Grade	Tumor stage	Familial
N1/T1	Luminal A	1	Ila	No
N2/T2	Luminal A	1	I	No
N3/T3	Basal	2	Ila	No
N4/T4	Luminal B	3	IIla	BRCA2 mut
N5/T5	TNP	3	Ila	BRCA2 mut
N6/T6	Basal	3	NA	No
N7/T7	Basal	NA	NA	No
N8/T8	Basal	NA	I	No
N9/T9	Basal	3	I	No
N10/T10	Luminal A	NA	IIb	No
N11/T11	Basal	2	Ila	BRCA2 mut
N12/T12	Basal	3	IIIb	No

Table S4. Primers used in bisulfite DNA sequencing and MS-HRM for the validation of methylation status in selected CpGDMs.

Probe	Gene	Region	N° evaluated CpGs	Primer	Sequence
Bisulfite sequencing					
cg26568226	<i>CYFIP1</i>	5'-UTR	13	BS_CYFIP1-F BS_CYFIP1-R	5' TTGAGAGGAGAATTTGAGAG 3' 5' ACTAAACACCAAACATAACCTC 3'
cg14024502	<i>MAP3K6</i>	promoter	8	BS_MAP3K6-F BS_MAP3K6-R	5' GTTTAGGGTGTAGGTTTTTTTT 3' 5' TAAAACTCAACCTCTCCCC 3'
cg01229567 cg19246761	<i>MIB2</i>	promoter	8	BS_MIB2-F BS_MIB2-R	5' TTGTAGGGAAAATTTTTAGGATT 3' 5' TATCAACCTAAAAAAAACCCA 3'
MS-HRM					
cg26568226	<i>CYFIP1</i>	5'-UTR	9	CYFIP1-F CYFIP1-R	5' GGAGAATTTGAGAGGTTGGT 3' 5' CTAACACCACCCACCTTATCC 3'
cg14024502	<i>MAP3K6</i>	promoter	4	MAP3K6-F MAP3K6-R	5' TTAGGGTTAGGGAAAAGGTT 3' 5'CCATCCCAATCCCTCCCTACCTA 3'
cg19246761 cg01229567	<i>MIB2</i>	promoter	4	MIB2-F MIB2-R	5' AGTATTTTAGATAAGTAGTTT 3' 5' CAAAAACCACTAACTATAAA 3'
cg09580608	<i>GNA13</i>	1st exon	10	GNA13-F GNA13-R	5' TTGATTAYGTTGTTGTAGATGG 3' 5' AAATCCACCTTCCTAAAACAA 3'
cg04890607	<i>HMGA2</i>	body gene	4	HMGA2-F HMGA2-R	5' TTTTAAAGTAGTAGGTGGTTT 3' 5' CCGAAAAATAATCTAACTCC 3'
cg24840062	<i>CDCP1</i>	body gene	2	CDCP1-F CDCP1-R	5' GGAATTTTGTGTTTATTGTTTT 3' 5'CTAACTCTTAAACCACTTTAA 3'

Table S5. Comparison of the genomic distribution of 38 CpGDMs to the expected distribution according to all printed CpG in the 450K HumanMethylation array. TSS1500: 1500 bp of transcription initiation site, TSS200: 200 bp of transcription initiation site, UTR: untranslated region.

Genomic region	% CpG on microarray	CpGDM (n)	Expected CpG (n)
TSS1500	14.97	9	5.7
TSS200	11.13	2	4.2
5'UTR	11.41	5	4.3
1st exon	6.83	1	2.6
Gene body	31.61	12	12
3'UTR	3.44	0	2.6
Intergenic	20.59	8	7.8

* Fisher test p value = 0.480

Table S6. CpGDMs overlapping genes previously associated to cancer (COSMIC and G2SBC Database).

Gene	Cancer type	Methylation direction	Genomic context	CpG island context
<i>AMOTL1</i>	breast	hypomethylation	TSS200	Island
<i>CDCP1</i>	breast	hypomethylation	gene body	open sea
<i>CYFIP1</i>	breast	hypermethylation	5'UTR	island
<i>GNA13</i>	lymphoma, gastric cancer	hypomethylation	1st Exon	island
<i>HMGA2</i>	lipoma, salivary gland adenoma	hypomethylation	gene body	open sea
<i>MAP3K6</i>	breast, ovary	hypomethylation	TSS1500	shore
<i>MIB2</i>	breast	hypomethylation	TSS1500	shore
<i>SDK1</i>	breast	hypomethylation	gene body	shore
<i>STK33</i>	pancreatic tumour	hypomethylation	TSS1500	island
<i>TAL1</i>	breast, ALL	hypomethylation	gene body	island
<i>TYROBP</i>	breast	hypomethylation	TSS1500	open sea

ALL: acute lymphoblastic leukemia, AML: acute myeloblastic leukemia, TSS200: 200 bp from transcription start site, TSS1500: 1500 bp from transcription start site.

Table S7. Overlapping CpGDMs detected in the blood of breast cancer patients from Latino population and in primary tissues from an independent European cohort of breast cancer patients.

CpG	CHR	UCSC REFGENE	Genomic context	CpG island context	P value (FDR)
cg04400047	1	<i>UBIAD1</i>	TSS1500	shore	0.007
cg06432479	1	<i>TAL1</i>	gene body	island	0.007
cg26874367	2		intergenic	open sea	0.021
cg25616514	3		intergenic	open sea	9.84 x10 ⁻⁹
cg01615258	3		intergenic	shore	1.12 x10 ⁻⁴
cg01311537	10	<i>C10orf128</i>	gene body	open sea	7.82 x10 ⁻⁴
cg13100962	11		intergenic	open sea	9.80 x10 ⁻⁹
cg23460961	11		intergenic	open sea	0.028
cg22313519	19	<i>KIAA1683</i>	TSS1500		7.79 x10 ⁻⁴

Table S8. *In silico* methylation analysis of selected candidate CpGDMs in breast primary tumors (n=735) and normal breast tissues (n=89) of TCGA database.

Gene	cg	Breast primary tumors Mean \pm SD	Breast normal tissues Mean \pm SD	p value*
<i>HMGA2</i>	cg04890607	0.597 \pm 0.249	0.713 \pm 0.144	2.8 x 10 ⁻⁴
<i>GNA13</i>	cg09580608	0.048 \pm 0.026	0.059 \pm 0.042	0.049
<i>MAP3K6</i>	cg14024502	0.531 \pm 0.127	0.527 \pm 0.054	0.735
<i>CYFIP1</i>	cg26568226	0.946 \pm 0.060	0.973 \pm 0.020	9.3 x 10 ⁻⁴
<i>CDCP1</i>	cg24840062	0.720 \pm 0.150	0.599 \pm 0.168	6.2 x 10 ⁻¹²

*Wilcoxon Rank sum test